

4 Working Groups

4.1 Report of Working Group on Ethics and Big Data

Bettina Berendt (KU Leuven, BEL), Geoffrey Rockwell (University of Alberta, CAN)

License © Creative Commons BY 3.0 Unported license
© Bettina Berendt, Geoffrey Rockwell

Joint work of Bettina Berendt, Chris Biemann, Marco Böhler, Geoffrey Rockwell, Joachim Scharloth, Claire Warwick

The following is the report of the Working Group on Ethics and Big Data (EBD) at the Dagstuhl seminar on Computer Science and Digital Humanities¹. This working group was formed to discuss ethical and privacy issues around big data following the Snowden revelations. Some of the questions we asked included:

- What are the ethical and privacy issues raised by big data methods?
- What are our responsibilities as researchers and educators working around big data?

We came to the conclusion that, whatever position one might take on the ethics of big data, we have responsibility to expose our students to the lively discussion around the issue. This led to a more focused question:

- How can we teach the ethics of big data?

During the course of our deliberations we did the following:

- We experimented with a close reading of the CSEC slides.² The idea was to use slides leaked by Snowden to both a) explore EBD across disciplinary boundaries and b) to experiment with a way of teaching EBD through current materials. Such close reading of primary source documents about big data and surveillance can bring CH and DH folk together. We need the CH folk to read the software represented and the DH folk to read the documents as rhetorical documents. There is an interesting opportunity also for joint research at this intersection.
- We discussed the literature and archives that need to be explored in this area. (See the Appendix below for some of the archives identified). We agreed to share resources. Rockwell has, for example, create a preliminary reading list to be built on.³
- We agreed to share pedagogical materials. Berendt has shared her materials and other plan to as they experiment with teaching EBD.⁴
- We discussed the development of an infographic that makes the case for the importance of ethics in big data.
- We agreed to develop a web site with resources on this subject. Böhler has set up the basic infrastructure for this and we will begin to populate it as we experiment with teaching EBD.
- We agreed to write a short (5000 word) opinion piece for the “Discussions” column of KI – Künstliche Intelligenz (<http://www.springer.com/computer/ai/journal/13218>). We outlined an argument we were all comfortable with as a way of developing a common ethic. (See Appendix A: Discussion Outline).

¹ See <http://www.dagstuhl.de/en/program/calendar/semhp/?semnr=14301>

² See http://www.scribd.com/fullscreen/188094600?access_key=key-2dvzky8d3gnowt96adba&allow_share=true&view_mode=scroll

³ See <http://philosophi.ca/pmwiki.php/Main/BigDataEthicsReadings>

⁴ See <http://people.cs.kuleuven.be/~bettina.berendt/teaching/2014-15-1stsemester/kaw/index1.htm> – these materials are used in a course described here <http://people.cs.kuleuven.be/~bettina.berendt/teaching/2014-15-1stsemester/kaw/>

Conclusion: It turns out that the issues are compelling and the reading of original documents like the CSEC slides to understand what the NSA (and others) are doing is one way into a shared discussion about EBD. Some of our conclusions were that:

- There are good ways to get people engaging with the issues – both the issues of ethics and the issues of what be really done get raised.
- We can imagine how we can turn forensic/diplomatic readings into problems for students and a site for interdisciplinary research.
- What the CSEC slides show is a process not unlike what we do ourselves (or want to do). This raises the issue of what the difference is between academic work and SIGINT (Signals Intelligence)? What makes one use of big data methods ethical or not?
- If the big data processes revealed by the Snowden leaks show good (or at least interesting) examples of big data interpretation (or analysis) then can we learn from them? Would it be ethical to copy the tools or processes revealed?
- Ultimately we have to ask how surveillance is different from research or forms of care for the other? Both are a form of knowing another – how is the other different and how is the knowing different?

Appendix A: Discussion Outline

- 1 Introduction (framing it in terms of the current discussion)
 - 1.1 How should we do big/data/science in light of the Snowden revelations?
 - 1.2 Background to Snowden revelations
- 2 What do academics have to offer? What is the role of the researcher now?
 - 2.1 We teach big data,
 - 2.2 We are researchers developing new methods and tools,
 - 2.3 We provide data,
 - 2.4 We are citizens, and
 - 2.5 We can act as mediators
- 3 The standard position on the ethics of big data is that it is not my business – that big data is just a tool/technique
 - 3.1 How do researchers talk about their developments?
 - 3.2 What can we learn from philosophy of technology?
 - 3.3 Mining is about discriminating – one cannot avoid legal and ethical issues
- 4 We should beware inventing the other – we need to ask about our activities in the academy too. How is our research a form of surveillance.
 - 5 Therefore we see the need for public dialogue rather than ethical decision trees that absolve people of the need to think about what they are doing
 - 5.1 Most people think research ethics boards are the solution – but this delegates
 - 5.2 We also need thick description – story telling
 - 5.3 Finally, we need to teach people across humanities/data sciences

Appendix B: Archives and Literature

- ACLU: <https://www.aclu.org/nsa-documents-search>
- Der Spiegel: <http://www.spiegel.de/international/the-germany-file-of-edward-snowden-documents-available-for-download-a-975917.html>
- Nymrod: <http://www.spiegel.de/media/media-34098.pdf>
- Cryptome: <http://www.cryptome.org>
- LeakSource: <http://leaksource.info/category/nsa-files/>